

METHOD OF FORMING DIFFERENT OXIDE THICKNESS FOR HIGH VOLTAGE TRANSISTOR AND MEMORY CELL TUNNEL DIELECTRIC

5

FIELD OF THE INVENTION

The present invention relates to a method of forming a semiconductor memory array of floating gate memory cells. The present invention also relates to a semiconductor memory array of floating gate memory cells of the foregoing type.

10

BACKGROUND OF THE INVENTION

Non-volatile semiconductor memory cells using a floating gate to store charges thereon and arrays of such non-volatile memory cells formed in a semiconductor substrate are well known in the art. Typically, such floating gate memory cells have been of the split gate type or stacked gate type. The memory cell utilizes a control gate that is spaced apart and insulated from a floating gate. The control gate can further include a select gate portion. A layer of insulating material (e.g. oxide) is formed between the floating gate and control gate. The thickness of this oxide layer is important because it dictates the level of capacitive coupling between the two gates and it must be thin enough to allow Fowler-Nordheim tunneling for those cells that are erased by tunneling electrons from the floating gate to the control gate.

20

Typically such memory cells are formed in an array configuration, with peripheral circuitry also formed on the same semiconductor die. The peripheral circuitry often includes one or more high voltage MOS transistors for operating the memory cell array. Fig. 1 illustrates a conventional memory cell 1 from the memory cell area of the die, as well a conventional high voltage MOS transistor 2, all formed on the same semiconductor die 3. The memory cell 1 is a split gate type memory cell. A tunneling oxide layer 4 is formed between the control gate CG and the floating gate FG, with source S and drain D regions formed in the substrate. The transistor 2 includes a poly gate PG that is insulated from the substrate 3 by a gate oxide 5, and overlaps with source S and drain D, where the poly gate PG controls the conductivity of the channel region between the source S and drain D.

30

Processing efficiency is important in the fabrication of semiconductor devices. Therefore, it is desirable to fabricate corresponding memory and transistor elements using the

same processing steps. Thus, the memory cell control gates CG and the transistor poly gates PG are preferably formed using the same polysilicon deposition step. Likewise, the formation of the memory cell tunnel oxide 4 and the transistor gate oxide 5 is formed using the same oxide formation step.

5 Figs. 2A to 2D illustrate a conventional method of forming a memory cell in the memory cell area 6 of substrate 3, and a MOS transistor in the peripheral area 7 of substrate 3. Silicon dioxide (hereinafter "oxide") layer 8, polysilicon (hereinafter "poly") layer 9, and silicon nitride (hereinafter "nitride") layer 10 are first formed over the substrate 3. A masking step is used to selectively etch and remove a portion of nitride layer 10, forming a
10 hole 11 that exposes the poly layer 9. The structure is oxidized to form oxide layer 12 over the exposed portion of poly layer 9 at the bottom of hole 11. The resulting structure is shown in Fig. 2A.

Next, a series of etches are used to remove nitride layer 10, and those portions of poly layer 9 and oxide layer 8 not protected by oxide layer 12, as shown in Fig. 2B. Oxide layer
15 13 is then formed over the structure, followed by the formation of nitride spacers 14, as shown in Fig. 2C. Finally, a poly deposition and masking step is used to form poly block 15 on oxide layer 13 (laterally adjacent and vertically over poly layer 9), and poly block 16 on oxide layer 13 (in peripheral area 7). Ion implantation is used to form source regions 17 and drain regions 18, as shown in Fig. 2D. Poly block 15 forms the memory cell control gate,
20 poly layer 9 forms the memory cell floating gate, and the portion of oxide layer 13 therebetween forms the memory cell tunnel oxide. Poly block 16 forms the transistor poly gate, and the portion of oxide layer 13 underneath poly block 16 forms the gate oxide for the transistor.

This fabrication process illustrates how the oxide layer 13 serves as both the tunnel
25 oxide for the memory cell 1 and the gate oxide for the transistor 2. Thus, the thickness of the tunnel and gate oxides is necessarily the same. However, this is not ideal for devices in which the MOS transistor 2 operates at a higher voltage than does the memory cell 1.

SUMMARY OF THE INVENTION

30 The present invention solves the aforementioned problem by providing memory cells and MOS transistors on the same silicon substrate, wherein the same oxide layer forms the

gate and tunnel oxides, yet the MOS transistor gate oxide thickness is greater than the thickness of the memory cell tunnel oxide.

5 The present invention is an electrically programmable and erasable memory device that includes a substrate of semiconductor material having a memory area and a peripheral area, a memory cell formed in the memory area of the substrate, and a MOS transistor formed in the peripheral area of the substrate. The memory cell includes an electrically conductive floating gate disposed over and insulated from the substrate, an electrically conductive control gate disposed adjacent to the floating gate, and an insulating layer formed in the memory and peripheral areas that includes a first portion that is disposed between the control gate and the floating gate with a thickness permitting Fowler-Nordheim tunneling of charges therethrough. The MOS transistor includes an electrically conductive poly gate disposed over and insulated from the substrate, and a second portion of the insulating layer being disposed between the poly gate and the substrate and having a thickness that is greater than that of the first portion of the insulating layer. The first and second portions of the insulating layer are initially formed as a continuous layer of material.

10 In another aspect of the present invention, the electrically programmable and erasable memory device includes a substrate of semiconductor material that includes a memory area and a peripheral area, a memory cell formed in the memory area of the substrate, and a MOS transistor formed in the peripheral area of the substrate. The memory cell includes a first source region and a first drain region formed in the substrate with a first channel region therebetween, an electrically conductive floating gate disposed over and insulated from at least a portion of the first channel region, an electrically conductive control gate disposed adjacent to the floating gate, and an insulating layer formed in the memory and peripheral areas, wherein the insulating layer has a first portion that is disposed between the control gate and the floating gate with a thickness permitting Fowler-Nordheim tunneling of charges therethrough. The MOS transistor includes a second source region and a second drain region formed in the substrate with a second channel region therebetween, an electrically conductive poly gate disposed over and insulated from at least a portion of the second channel region, and a second portion of the insulating layer being disposed between the poly gate and the second channel region, wherein the second portion of the insulating layer has a thickness that

is greater than that of the first portion of the insulating layer. The first and second portions of the insulating layer are initially formed as a continuous layer of material.

In yet another aspect of the present invention, a method of making an electrically programmable and erasable memory device includes the steps of forming an electrically
5 conductive floating gate disposed over and insulated from a memory area of a substrate, forming an insulating layer that has a first portion formed over the memory area of the substrate and a second portion formed over a peripheral area of the substrate, wherein the insulating layer first portion has a thickness permitting Fowler-Nordheim tunneling of charges therethrough, changing a thickness of one of the insulating layer first and second
10 portions relative to the other of the insulating layer first and second portions, forming an electrically conductive control gate disposed adjacent to the floating gate and insulated therefrom by the first portion of the insulating layer, and forming an electrically conductive poly gate disposed over the peripheral area of the substrate and insulated therefrom by the second portion of the insulating layer.

In yet one more aspect of the present invention, the method of forming an electrically
15 programmable and erasable memory device includes the steps of forming a memory cell in a memory area of a substrate, and forming an MOS transistor in a peripheral area of the substrate. The memory cell formation includes the steps of forming a floating gate over and insulated from the substrate, and forming a control gate adjacent to and insulated from the
20 floating gate. The MOS transistor formation includes forming a poly gate over and insulated from the substrate. The formation of the memory cell and the formation of the MOS transistor together include the step of forming an insulating layer having a first portion that is disposed between the control gate and the floating gate with a thickness permitting Fowler-Nordheim tunneling of charges therethrough, and a second portion that is disposed between
25 the poly gate and the substrate, wherein the second portion of the insulating layer has a thickness that is greater than that of the first portion of the insulating layer.

Other objects and features of the present invention will become apparent by a review of the specification, claims and appended figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a cross sectional view of a conventional memory cell and related peripheral MOS transistor.

5 Figs. 2A-2D are cross sectional views of a conventional semiconductor structure showing in sequence the steps in the processing of the semiconductor structure in the formation of the conventional memory cell and peripheral MOS transistor of Fig. 1.

Figs. 3A-3G are cross sectional views of a semiconductor structure showing in sequence the steps in the processing of the semiconductor structure in the formation of the memory cell and peripheral MOS transistor of the present invention.

10 Figs. 4A-4C are cross sectional views of a semiconductor structure showing in sequence the steps in an alternate embodiment of the processing of the semiconductor structure shown in Fig. 3C.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

15 The method of the present invention is illustrated in Figs. 3A-3G, and begins with a semiconductor substrate 30, which is preferably of P type and is well known in the art. The thicknesses of the layers described below will depend upon the design rules and the process technology generation. It will be understood by those skilled in the art that the present invention is not limited to any specific process technology generation, nor to any specific value in any of the process parameters described hereinafter. In addition, the following method description focuses on the formation of a single memory cell and a single MOS transistor, when in fact an array of such memory cells is simultaneously formed in a memory cell area 32 of the substrate 30, and a plurality of such transistors can be simultaneously formed in a peripheral area 34 of the substrate 30. Preferably, such memory cells are formed in columns of active regions that are separated by columns of insulation areas, the formation of which is well known in the art.

25 As shown in Fig. 3A, a thin layer of insulation material 36, such as oxide having a thickness approximately $\sim 50\text{-}150 \text{ \AA}$, is formed on substrate 30 by any well known technique such as oxidation or deposition (e.g. chemical vapor deposition or CVD). A layer of 30 conductive material 38, such as polysilicon having a thickness of approximately $\sim 500\text{-}2000 \text{ \AA}$ is deposited on top of the oxide layer 36 by any well known process such as Low Pressure

CVD or LPCVD. Poly layer 38 may be doped by ion implantation. Another layer of insulation material 40, such as nitride having a thickness of approximately $\sim 300\text{-}500\text{ \AA}$, is deposited over the poly layer 38, preferably by CVD. Layers 36/38/40 are formed in both the memory cell area 32 and the peripheral area 34 of the semiconductor substrate 30. Next, a conventional photo-lithography scheme is used to form semi-recessed holes or trenches 42 in the structure in the following manner. A suitable photo-resistant material is applied on the nitride layer 40 and a masking step is performed to selectively remove the photo-resistant material from a selected area. Where the photo-resist material is removed, the exposed portion of nitride layer 40 is removed using a standard nitride etch process (e.g. anisotropic nitride etch), leaving a hole or trench 42 that exposes the poly layer 38. The structure is further processed to remove the remaining photo resist. A thermal oxidation process is used to oxidize the exposed portion of poly layer 38 in hole 42, to form a lens-shaped oxide layer 44 over the exposed portion of poly layer 38, as shown in Fig. 3A.

Next, a nitride etch process is used to remove the remaining portions of nitride layer 40. This is followed by anisotropic poly and oxide etch steps to remove those portions of poly layer 38 and oxide layer 36 that are not protected by oxide layer 44. The resulting structure is shown in Fig. 3B.

Oxide layer 46 is formed over the structure, preferably using a combination of a high temperature oxide deposition process (HTO) with a thickness of approximately $100\text{-}200\text{ \AA}$, and a thermal oxidation process with a thickness of $10\text{-}50\text{ \AA}$. A layer of nitride 48 is formed over the oxide layer 46, having a thickness of approximately $\sim 100\text{-}300\text{ \AA}$. This is followed by the formation of photo resist material over the structure, and then a masking step in which the photo resist material is removed from the peripheral area 34 only, leaving a layer of photo resist material 50 over the memory cell area 32, as shown in Fig. 3C.

A nitride etch process is used to remove the nitride layer 48 from the peripheral area 34, while nitride layer 48 in the memory cell area 32 (protected by photo resist 50) remains intact. The photo resist 50 is then removed to result in the structure shown in Fig. 3D.

A thermal oxidation is next used to thicken oxide layer 46 in the peripheral area 34, but not in the memory cell area 32. Nitride layer 48 in the memory cell area 32 acts as a mask so that the thermal oxidation process only grows thicker the exposed portion of oxide layer 46 in the peripheral area 34. The resulting structure is shown in Fig. 3E.

An anisotropic nitride etch is used to remove nitride layer 48, except for nitride spacers 52 formed on either side of the raised portion (over oxide layer 44) of oxide layer 46, as shown in Fig. 3F. Polysilicon is deposited on oxide layer 46, and a masking step is used to remove selected portions of the deposited polysilicon, except for poly block 54 in the memory area 32 and for poly block 56 in the peripheral area 34. In the embodiment shown in Fig. 3G, the poly block 54 has a first portion disposed laterally adjacent to poly layer 38, and a second portion disposed vertically over poly layer 38, with oxide layer 46 disposed between poly layer 38 and poly block 54. Conventional ion implantation is then made across the structure to form source and drain regions 58 and 60 in the memory area 32 of the substrate 30, and source and drain regions 62 and 64 in the peripheral area 34 of the substrate 30. The source and drain regions have a conductivity type that is different from that of the substrate or substrate well in which they are formed. In the preferred embodiment, the source and drain regions are all N⁺ regions. The resulting structure is shown in Fig. 3G.

In the memory cell area 32, a channel region 66 is defined in the substrate between the source 58 and drain 60. Poly layer 38 forms the cell's floating gate, which is disposed over and insulated from a first portion of the channel region 66 and a portion of the source region 58. Poly block 54 forms the cell's control gate, which includes a first portion 54a that is disposed over and insulated from a second portion of the channel region 66 and a portion of the drain 60, and is laterally adjacent to and insulated from the floating gate 38. The control gate 54 has a second portion 54b that is disposed over (vertically adjacent to) and insulated from the floating gate 54. A notch 68 is formed in the control gate 54 by the nitride spacer 52, which helps prevent reverse tunneling back to the floating gate or to the substrate. The non-volatile memory cell is of the split gate type as described in U.S. Patent No. 5,572,054, which discloses the operation of such a non-volatile memory cell and an array formed thereby, and is hereby incorporated herein by reference.

In the peripheral area 34, a channel region 70 is defined in the substrate between the source 62 and drain 64. Poly block 56 forms the transistor cell's poly gate, which is disposed over and insulated from the channel region 70.

Oxide layer 46 has two separate portions: a thinner portion 72 formed in the memory cell area 32 of the substrate 30, and a thicker portion 74 formed in the peripheral area 34 of the substrate 30. Part of the thinner portion 72 of oxide layer 46 forms the tunnel oxide of

the memory cell through which electrons tunnel for memory cell operation (e.g. erase operation). Part of the thicker portion 74 of the oxide layer 46 forms the gate oxide of the MOS transistor which insulates the poly gate 56 from the substrate 70. Because the gate oxide of the MOS transistor is formed on a thicker portion of oxide layer 46 than is the
5 tunneling oxide of the memory cell, the MOS transistor can better operate at a higher voltage than that of the memory cell. In the preferred embodiment, the control gate is raised to an erase voltage of approximately 7 to 15 volts, and the MOS transistor operates at 1 or 2 volts above that erase voltage. This difference of 1-2 volts provides a margin for better reliability in the operation of the semiconductor memory device.

10 Figures 3A-3G show the formation of oxide layer 46 over both the peripheral area 34 and the memory cell area 32, and then thickening of the oxide layer 46 only in the peripheral area 34. Alternately, however, the thickness of the oxide layer 46 in the memory cell area can be reduced, resulting in the same relative thicknesses of oxide layer 46 in the peripheral and memory cell areas 34/32. Figures 4A-4C illustrate an alternate embodiment of the
15 present invention, which essentially involves the same processing steps as previously described with respect to Figs. 3A-3G, but with the following exceptions. This alternate embodiment begins with the same structure as shown in Fig. 3C, except that the thickness of oxide layer 46 is greater, and the photo resist 50 is formed over the peripheral area 34 instead of memory cell area 32, as shown in Fig. 4A.

20 The subsequent nitride etch process removes the nitride layer 48 from the memory cell area 32, while nitride layer 48 in the peripheral area 34 (protected by photo resist 50) remains intact. The photo resist 50 is then removed to result in the structure shown in Fig. 4B.

An oxide etch process is then used to reduce the thickness of oxide layer 46 in the
25 memory cell area 34, but not in the peripheral area 34. The nitride layer 48 in the peripheral area 34 acts as a mask so that the oxide etch process only removes the top exposed portion of oxide layer 46 in the memory cell area 32. The resulting structure is shown in Fig. 4C, which is further process as described above with respect to Fig. 3G. It should be noted that nitride layer 48 is optional in this embodiment, and that a nitride deposition and removal step can be
30 used to add the nitride spacer 52 to prevent reverse tunneling back to the floating gate or to the substrate in the finished memory cell.

It is to be understood that the present invention is not limited to the embodiment described above and illustrated herein, but encompasses any and all variations falling within the scope of the appended claims. For example, the invention is disclosed in the context of a split gate memory device, but it is also applicable to any memory cell design that includes a tunneling oxide between two conductive elements, where the tunneling oxide layer extends to the substrate's peripheral area to form the gate oxide of a MOS transistor. Further, although the foregoing method describes the use of appropriately doped polysilicon as the conductive material used to form the memory cells, it should be clear to those having ordinary skill in the art that any appropriate conductive material can be used. In addition, any appropriate insulator can be used in place of silicon dioxide or silicon nitride. Moreover, any appropriate material whose etch property differs from silicon dioxide (or any insulator) and from polysilicon (or any conductor) can be used in place of silicon nitride. Further, as is apparent from the claims, not all method steps need be performed in the exact order illustrated or recited in the claims, but rather can be performed in any order that allows the proper formation of the memory cell and MOS transistor of the present invention. Finally, while oxide layer 46 is formed over the both the peripheral and memory cell areas 34/32 as a single layer of material, this oxide layer could in fact be formed of multiple layers of insulating material, and could later be made discontinuous by the formation of holes or trenches therein.